Osteoarthritis and Cartilage 20 (2012) 548–556

Osteoarthritis and Cartilage

Evaluation of separate quantitative radiographic features adds to the prediction of incident radiographic osteoarthritis in individuals with recent onset of knee pain: 5-year follow-up in the CHECK cohort

M.B. Kinds †‡, A.C.A. Marijnissen †, K.L. Vincken ‡, M.A. Viergever ‡, K.W. Drossaers-Bakker §, J.W.J. Bijlsma †, S.M.A. Bierma-Zeinstra ||¶, P.M.J. Welsing †#, F.P.J.G. Lafeber †*

† *Rheumatology & Clinical Immunology, University Medical Center Utrecht, the Netherlands*
‡ *Image Sciences Institute, University Medical Center Utrecht, the Netherlands*
§ *Arthritis Center Twente, University Twente & Medisch Spectrum Twente, Enschede, the Netherlands*
|| *General Practice, University Medical Center Rotterdam, the Netherlands*
¶ *Orthopaedics, University Medical Center Rotterdam, the Netherlands*
# *Julius Center for Health Sciences & Primary Care, University Medical Center Utrecht, the Netherlands*

ARTICLE INFO

Article history:
Received 6 September 2011
Accepted 16 February 2012

Keywords:
Osteoarthritis
Knee
Early
Development
Radiographic
Features

SUMMARY

Objective: Detailed radiographic evaluation might enable the identification of osteoarthritis (OA) earlier in the disease. This study evaluated whether and which separate quantitative features on knee radiographs of individuals with recent onset knee pain are associated with incidence of radiographic OA and persistence and/or progression of clinical OA during 5-year follow-up.

Method: From the Cohort Hip & Cohort Knee study participants with knee pain at baseline were evaluated. Radiographic OA development was defined as Kellgren & Lawrence (K&L) grade ≥II at 5-year follow-up. Clinical OA was defined as persistent knee pain and as progression of Westen Ontario & McMaster Universities Osteoarthritis index (WOMAC) pain and function score during follow-up. At baseline radiographic damage was determined by quantitative measurement of separate features using Knee Images Digital Analysis, and by K&L-grading.

Results: Measuring osteophyte area [odds ratio (OR) = 7.0] and minimum joint space width (OR = 0.7), in addition to demographic and clinical characteristics, improved the prediction of radiographic OA 5 years later [area under curve receiver operating characteristic = 0.74 vs 0.64 without radiographic features]. When the predictive score (based on multivariate regression coefficients) was larger than the cut-off for optimal specificity, the chance of incident radiographic OA was 54% instead of the prior probability of 19%. Evaluating separate quantitative features performed slightly better than K&L-grading (AUC = 0.70). Radiographic characteristics hardly added to prediction of clinical OA.

Conclusion: In individuals with onset knee pain, radiographic characteristics added to the prediction of radiographic OA development 5 years later. Quantitative radiographic evaluation in individuals with suspected OA is worthwhile when determining treatment strategies and designing clinical trials.

Introduction

Osteoarthritis (OA) is a disabling joint disease which most commonly affects the knee joint. Symptoms of pain and functional limitations are assumed to be originated by structural changes like articular cartilage damage, osteophyte formation, synovial inflammation, and subchondral bone changes[1,2]. Radiography is the gold standard for demonstrating structural changes since image acquisition is non-invasive, cheap, fast, and generally available[1,3]. Radiographic OA is commonly graded according to Kellgren & Lawrence (K&L)[4]. A drawback is that this method only provides a qualitative (ordinal) score of a combination of structural aspects. It is generally appreciated that it takes at least a year before a change of one grade (scale 0–IV) becomes evident[5,6]. More detailed evaluation by quantitative measurement of separate features of joint damage might improve the association with clinical symptoms, which is currently not consistently found[7,8]. More importantly, it might enable the identification of initial tissue

* Address correspondence and reprint requests to: F.P.J.G. Lafeber, University Medical Center Utrecht, Rheumatology & Clinical Immunology, F02.127, PO Box 85500, 3508 GA Utrecht, the Netherlands. Tel: 31-88-75-58521; Fax: 31-30-252-3741.
E-mail address: F.Lafeber@umcutrecht.nl (F.P.J.G. Lafeber).

1063-4584/$ – see front matter © 2012 Osteoarthritis Research Society International. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.joca.2012.02.009

damage earlier in the disease, which is of value for the development of interventions to prevent structural damage[9].

In clinical practice patients visit a physician when they suffer from complaints that are possibly related to OA. Early in the disease process reliable diagnosis is difficult because structural damage can not yet be detected on radiographs using methods like K&L-grading. Also, pain often has an intermittent character and not all individuals suspected for the disease will eventually develop progressive OA. Higher age[10], higher body mass index (BMI)[11,12], and female gender[10] have been shown to be associated with the onset and progression of OA. In addition, the detection of early evident tissue damage by precise measurement on radiographs might predict the radiographic and/or clinical course of disease[13,14]. In the hip joint the measurement of smaller joint space width (JSW) was found to predict total hip replacement[15,16], but in the knee joint initial severity has not been found to be of evident additional value in the prediction of radiographic[17] and clinical progression[18].

The objective of the present study was to evaluate whether and which separate features of radiographic damage, measured quantitatively in knees with early symptoms related to OA, are associated with the incidence of radiographic OA and the persistence and/or progression of clinical OA during 5-year follow-up.

Methods

Cohort Hip & Cohort Knee (CHECK)

Cohort Hip & Cohort Knee (CHECK) is a Dutch prospective multicentre 10-year follow-up study. Individuals ($n = 1002$) with pain and/or stiffness of hip and/or knee, age 45–65 years, and without a previous visit or with a first visit no longer than 6 months ago to the general practitioner for these complaints were included. Individuals with pathological conditions other than OA explaining the complaints and individuals with K&L-grade IV were excluded[19]. The study procedures are in accordance with the standards of the medical ethics committees of all 10 participating hospitals and with the Helsinki Declaration of 1975 (as revised in 2000), and all participants gave their written informed consent.

Radiographic features baseline

Of both knees separately, standardized radiographs were acquired[20,21]. The baseline (T0) radiographs of both knees were evaluated for their predictive ability.

Radiographic parameters of knee OA were quantitatively measured by use of Knee Images Digital Analysis (KIDA)[22]. Key radiographic features were defined for evaluation in the present study, based on principal component analysis and on expert opinion (AM/JB/FL). The 'minimum JSW' (in mm) was measured as the smallest distance between femur and tibia. Also 'medial JSW' and 'lateral JSW' were determined by calculating the mean of four predefined locations (standardized, based on the joint dimensions[22]). The angle between the femur and tibia in the frontal plane was determined to represent the alignment of the joint ('varus angle' in degrees; positive value represents varus alignment). 'Osteophyte area' (in mm^2) was determined by summing the osteophyte area of the lateral and medial femur and tibia. 'Eminence height' was calculated as the sum of both eminences. 'Bone density' was determined as the mean of the bone density determined in the lateral and medial femur and tibia. Bone density was expressed in mmAl equivalents, by normalizing the grey values of the subchondral bone region to those of an aluminium reference wedge present on all radiographs[22]. The KIDA method is a mathematical interactive software tool to analyze knee radiographs and takes a few minutes per knee joint. Measurements were performed by one experienced observer (ML) in random order and blinded to any information (e.g., clinical characteristics). The intra-observer variation tested by random reanalysis of 108 radiographs several months later, revealed good intra-observer variability [intraclass correlation coefficient (ICC) = 0.73–0.99] for the different features.

The number of analyzed knees varied for the different radiographic features since KIDA measurement requires good radiographic quality. E.g., measurement of varus angle and eminence height was hampered in 10 (0.5% of 2004) knees, and osteophyte area could not be thoroughly outlined in 31 knees (1.5%). Specifically bone density measurement, which requires good contrast and a clearly visible aluminium reference wedge, was not always possible despite standardized procedures (28%). The baseline characteristics were not significantly different between participants without and with missing quality radiographs.

Radiographic OA at T0 was also assessed by the commonly used K&L-grading. K&L-grades were determined without knowledge of any other characteristics and reading was performed in pairs [T0 and 2-year follow-up (T2y)] to obtain a reliable grade for T0 (T2y data not used in the present study).

OA development

The development of OA from T0 to 5-year follow-up (T5y) was classified as 'poor' outcome (incidence, persistence, or progression) or 'good' outcome (no incidence, persistence, or progression) based on radiographic and clinical evaluation. OA development was evaluated in participants with complaints in at least one knee at study inclusion. For the different definitions of OA development separate analyses were performed with specific criteria for joint and participant exclusion (Table I).

Radiographic OA: incidence

The incidence of radiographic OA ('poor' outcome) was defined on joint level (knees separately), as the development of a K&L-grade ≥II at T5y. Since knees needed to be susceptible for the development of radiographic OA, knees with K&L-grade ≥II at T0 were excluded for these analyses. For each knee the T5y radiograph was graded according to K&L with the radiographs of T0 and T2y in view. The T5y radiograph was scored in another scoring session than the initial T0 grade, which was determined independently of the T5y outcome to prevent information bias.

Table I
Definitions of OA development used as outcome in the different analyses

	Radiographic OA incidence K&L	Clinical OA persistent pain	WOMAC pain	WOMAC function
Inclusion of participants/knees				
Criteria: at T0 painful knee, and… (*n*)	K&L <II (985 knees)	Only painful knee (1060 knees)	Hip: not painful and K&L <II (286 participants)	Hip: not painful and K&L <II (279 participants)
'Poor' outcome				
Definition	K&L ≥II at T5y	Painful at T4y&T5y	Quintiles: highest 3/move higher at T5y	Quintiles: highest 3/move higher at T5y
% of participants/knees	19%	48%	54%	56%

Clinical OA: persistence or progression

The development of 'poor' clinical outcome was evaluated on joint level and on participant level. For the definition on joint level all knee joints that were painful at T0 were used. The physician assessed this during examination of joint motion, for the left and right knee separately. Clinical persistence ('poor' outcome) was defined as still having a painful knee during physical examination both at 4-year and at 5-year follow-up (T4y and T5y), and otherwise the outcome was considered 'good'.

On participant level, the Western Ontario & McMaster Universities Osteoarthritis index (WOMAC) pain score and function score were used. For these analyses, confounding of the WOMAC scores by hip involvement was prevented by excluding participants with additional painful hip(s) and/or K&L-grade ≥II of the hip(s) at T0. WOMAC scores were standardized to a 0–100 scale with the maximum score representing the worst condition. Because WOMAC scores are recognized to be quite variable over time[23], the WOMAC 'baseline' value was calculated as the mean of T0 and T1y, and the 'follow-up' score was calculated as the mean of T4y and T5y. Development (and persistence) of WOMAC pain and function values from 'baseline' to 'follow-up' was classified according to *Sharma et al.* using a quintile approach[24]. The clinical progression was defined as 'poor' when participants moved to a higher quintile or remained in the highest three quintiles, and the outcome was defined as 'good' when participants moved to a lower quintile or remained in the lowest two quintiles.

Statistical analyses

Separate binary logistic regression analyses were performed with 'good' (0) *versus* 'poor' (1) radiographic or clinical outcome as dependent variable. The radiographic characteristics defined as separate key features (KIDA) or K&L-grade were used as independent variables [for osteophyte area 'log (osteophyte area + 1)' was calculated to obtain a more normal distribution]. For analyses on joint level (K&L and persistent pain outcome) the value of a knee was used, and for analyses on participant level (WOMAC outcome) the sum of the left and right knee was evaluated to represent the total burden of disease.

Also, since radiographic characteristics might be dependent on characteristics of an individual (e.g., larger individuals have larger JSW and females have lower bone density), the difference-value (difference between knee and contralateral knee) was used as independent variable as well[25].

Furthermore, gender, age, BMI, and erythrocyte sedimentation rate (ESR in mm/h), were added as independent variables. The latter was included because this parameter is frequently determined in this early stage of OA to exclude arthritic conditions. Also, dependent on the clinical outcome (persistence or progression), overall pain intensity (0–10 scale), WOMAC pain score, or WOMAC function score at baseline was added as independent variable.

Univariate and multivariate regression analyses were performed. In the multivariate analyses all variables were initially included and variables were removed manually using a backward stepwise selection procedure. Variables that were either statistically significantly ($P \leq 0.05$) related to the outcome or that changed the regression coefficient for one of the radiographic characteristics with >10% (confounding variables) were kept in the final model.

The approach aimed at representing clinical practice when a patient visits a physician with the first OA related symptoms analyzing whether radiographic characteristics (separate features or K&L-grade), next to the assessment of basic demographic and clinical characteristics, add to decision-making.

To evaluate the fit of the final models, Hosmer–Lemeshow tests were performed. Prognostic ability of the models was summarized using the area under curve (AUC) of the receiver operating characteristic (ROC). The AUC–ROC provides a measure for the ability to discriminate between 'good' and 'poor' outcome, where an AUC–ROC <0.70 was regarded as poor, 0.70–0.80 as fair, 0.80–0.90 as good, and ≥0.90 as excellent[26].

When the discriminative ability of the models was considered fair to good, also the sensitivity, specificity, and positive and negative predictive values (PPV and NPV) were calculated as prognostic statistics for different cut-off values. Therefore, the regression coefficients were corrected for over-fitting using the method of van Houwelingen and LeCessie[27] and the regression function was converted into a simple predictive score. Analyses were performed using SPSS 15.0, *P*-value ≤0.05 was considered statistically significant.

Results

Baseline characteristics

Of 1002 CHECK participants, 829 had pain in the knee(s) at study inclusion (1294 painful knees). Radiographs of good quality were available of 1060 (of the 1294) knee joints for analyses of persistent pain. Of these, data of 985 knees with K&L-grade <II at T0 were available for analyses of radiographic incidence. For analyses of WOMAC progression (participant level), data were available of 286 (WOMAC pain) and 279 (WOMAC function) participants with no hip affection at T0.

Table II depicts demographic and clinical as well as radiographic characteristics at T0 of the respective datasets, separately for the participants with 'good' and 'poor' radiographic and clinical outcome. Incidence of radiographic OA was observed in 189 of 985 knees (19% 'poor' outcome: K&L-grade ≥II at T5y). Persistent knee pain was observed in 509 of 1060 knees (48% 'poor' outcome). Clinical progression ('poor' outcome according to quintile approach) was observed in 155 of 286 (54%) and 155 (56%) of 279 participants evaluated for WOMAC pain and WOMAC function, respectively.

Predictors OA development

Radiographic OA: incidence

Table III depicts results of univariate and multivariate regression analyses with incidence of radiographic OA as dependent variable (K&L-grade ≥II at T5y). Odds ratios (ORs) with 95% confidence interval (95%CI) and statistical significance (*P*) are depicted for the independent variables determined at T0. For the multivariate models the AUC–ROC is depicted.

For the multivariate models the fit was adequate (Hosmer–Lemeshow tests: $P > 0.05$). The prognostic ability was clearly improved when radiographic characteristics were added to demographic and clinical characteristics. For the multivariate models using demographic and clinical characteristics, separate features, and K&L-grading as independent variables, data of respectively 965, 904, and 955 (of 985) knees were available. The model with basic demographic and clinical variables only revealed that female gender and higher BMI were statistically significant predictors of incidence of radiographic OA, which is in accordance with the literature[10–12]. Prognostic ability of this model was considered poor[26]: AUC–ROC = 0.64 (95%CI: 0.59–0.68, $P < 0.0001$). When key radiographic features obtained at T0 were added to the model the ability to predict radiographic outcome at T5y improved (AUC–ROC = 0.74, 95%CI: 0.69–0.78, $P < 0.0001$), resulting in fair prognostic ability. This AUC–ROC was statistically significantly higher than the AUC–ROC of the model with demographics and clinical variables only ($P = 0.007$), as evaluated according to Hanley and McNeil[28] (in participants with complete

Table II
Demographic, clinical and radiographic characteristics at T0

	K&L (joint)		Persistent pain (joint)		WOMAC pain (pp)		WOMAC function (pp)	
	'Good'	'Poor'	'Good'	'Poor'	'Good'	'Poor'	'Good'	'Poor'
n participants (k)	520 (796)	133 (189)	346 (551)	334 (509)	131	155	124	155
Age in years	56 (5)	56 (5)	56 (5)	56 (5)	56 (5)	56 (5)	56 (5)	56 (5)
Female gender	78%	86%	82%	80%	75%	81%	**73%**	**83%**
BMI in kg/m^2	**25 [23–28]**	**27 [24–31]**	25 [23–28]	26 [24–29]	**25 [23–27]**	**26 [24–29]**	**25 [23–27]**	**26 [24–29]**
ESR in mm/hour	8 [5–14]	8 [5–14]	8 [5–14]	8 [5–14]	**7 [4–11]**	**8 [5–14]**	**6 [4–11]**	**9 [5–15]**
Pain intensity	**3 [2–5]**	**4 [2–5]**	**3 [2–5]**	**4 [2–5]**	**3 [2–4]**	**3 [2–5]**	3 [2–5]	4 [2–5]
WOMAC pain	**20 [10–35]**	**25 [15–40]**	**20 [10–35]**	**25 [15–40]**	**20 [9–30]**	**20 [10–35]**	**15 [10–30]**	**20 [10–31]**
WOMAC function	**19 [9–34]**	**24 [13–38]**	**18 [10–31]**	**25 [13–38]**	**13 [6–25]**	**21 [10–32]**	**13 [6–27]**	**20 [10–32]**
K&L-grade ≥II	0%	0%	2.5%	4.6%	3.2%	4.2%	3.0%	4.5%
Radiographic features								
Minimum JSW	**3.07 (1.13)**	**2.65 (1.41)**	**3.04 (1.19)**	**2.85 (1.28)**	2.99 (1.25)	2.99 (1.26)	**3.14 (1.24)**	**2.84 (1.25)**
Medial JSW	**4.80 (0.87)**	**4.49 (1.12)**	4.74 (0.90)	4.65 (1.02)	4.59 (1.04)	4.75 (1.08)	4.79 (1.09)	4.57 (1.03)
Lateral JSW	6.07 (1.37)	6.24 (1.48)	6.09 (1.33)	6.01 (1.41)	6.10 (1.36)	6.13 (1.39)	6.15 (1.26)	6.11 (1.46)
Varus angle	**1.68 (1.70)**	**2.32 (2.03)**	1.78 (1.77)	1.79 (1.83)	1.98 (1.89)	1.83 (1.97)	1.81 (1.90)	2.02 (1.94)
Osteophyte	**5.20 (4.85)**	**10.21 (9.09)**	**5.76 (5.58)**	**7.24 (7.60)**	6.97 (7.31)	7.55 (6.85)	6.83 (6.20)	7.75 (7.69)
Eminence	**22.7 (3.1)**	**23.2 (3.1)**	22.8 (3.0)	22.8 (3.2)	22.6 (3.1)	22.8 (3.3)	22.7 (3.1)	22.7 (3.2)
Bone density	**24.5 (6.1)**	**26.1 (6.2)**	24.5 (5.6)	25.0 (6.4)	25.2 (6.5)	24.3 (6.4)	25.4 (6.3)	24.2 (6.5)

(Joint): defined at joint level, (pp): defined at participant level, (k): *n* knees, mean (standard deviation) or median [25–75th percentile] depicted, **bold**: significant difference between participants with 'good' and 'poor' outcome.

data for both models). This model implied that knees with *smaller* minimum JSW (OR = 0.74) and those with *larger* osteophyte area (OR = 6.97) were more likely to have incident radiographic OA ('poor' outcome). Also the K&L-grade (0 or I) at T0 added to clinical and demographic variables as a predictor for incidence of radiographic OA at T5y (OR = 4.74). This model had (borderline) fair prognostic ability with AUC–ROC of 0.70 (95%CI: 0.66–0.74, $P < 0.0001$), not statistically significantly different from the model with demographics only.

To evaluate whether the quantitative measurement of radiographic features can be applied in clinical practice, to identify individuals that are more likely to develop radiographic OA 5 years later, a simplified predictive score was calculated. The predictive score was based on the shrunken (shrinkage factor 0.98) and rounded regression coefficients (not the ORs as presented in the table) of the final logistic regression model including the key features of radiographic damage and demographic and clinical variables as: $-0.5 \times$ minimum JSW $+ 2 \times$ [log (osteophyte

Table III
Regression analyses with radiographic OA (K&L-grade ≥II) as dependent variable

Univariate				Multivariate				
	OR	(95%CI)	*P*		OR	(95%CI)	*P*	AUC–ROC
Demographic & clinical				**Demographic & clinical**				0.64*
Age	1.03	(1.00–1.06)	0.10					
Female gender	1.73	(1.10–2.72)	0.02	Female gender	1.68	(1.06–2.66)	0.03	
BMI	1.10	(1.06–1.14)	<0.0001	BMI	1.10	(1.06–1.14)	<0.0001	
ESR	1.00	(0.98–1.02)	0.76					
Pain intensity	1.08	(1.00–1.16)	0.05					
Radiographic				**Radiographic feature (demographic & clinical)**				
Key feature				***Key feature***				0.74*
Minimum JSW	0.75	(0.66–0.86)	<0.0001	Minimum JSW	0.74	(0.64–0.85)	<0.0001	
Medial JSW	0.67	(0.57–0.83)	<0.0001					
Lateral JSW	1.09	(0.97–1.22)	0.15					
Varus angle	1.22	(1.11–1.33)	<0.0001					
Osteophyte	6.30	(3.82–10.4)	<0.0001	Osteophyte	6.97	(4.11–11.8)	<0.0001	
Eminence	1.06	(1.00–1.12)	0.03					
Bone density	1.04	(1.01–1.08)	0.02					
Key feature diff								
Minimum JSW	0.90	(0.74–7.08)	0.25					
Medial JSW	0.97	(0.84–1.12)	0.67					
Lateral JSW	0.86	(0.66–1.11)	0.86					
Varus angle	1.03	(0.94–1.14)	0.59					
Osteophyte	1.83	(1.17–2.86)	0.009					
Eminence	0.98	(0.91–1.07)	0.70					
Bone density	1.01	(0.92–1.11)	0.76					
				Female gender	1.99	(1.20–3.31)	0.008	
				BMI	1.09	(1.05–1.14)	<0.0001	
				K&L (demographic & clinical)				
K&L-grade	3.54	(2.46–5.07)	<0.0001	***K&L-grade***	4.74	(3.13–7.19)	<0.0001	0.70
K&L diff	0.86	(0.62–1.20)	0.38	***K&L diff***	0.56	(0.40–0.78)	0.001	
				Female gender	1.90	(1.17–3.08)	0.009	
				BMI	1.09	(1.05–1.13)	<0.0001	

P: significance level; AUC–ROC: area under the receiver operating characteristic curve; diff: difference-value, note: female gender and BMI depicted for three multivariate models: demographic & clinical, and key feature and K&L-grade in addition to demographic & clinical; *statistically significantly different.

Table IV
Ability to predict radiographic incidence for three cut-off points of predictive score

Cut-off	2.50	3.65	4.60
Sensitivity	93%	66%	38%
Specificity	23%	66%	92%
PPV	23%	32%	54%
NPV	93%	89%	86%

PPV: positive predictive value, NPV: negative predictive value.

area + 1)] + 0.5 × gender + 0.1 × BMI. Based on the ROC-curve three cut-off points were evaluated for predictive ability: predictive score >2.50 (optimal sensitivity), score >3.65 (optimal trade-off between sensitivity and specificity), and score >4.60 (optimal specificity). Table IV shows sensitivity, specificity, PPV, and NPV for these cut-offs.

Using the predictive score the AUC–ROC was 0.73 (95%CI: 0.69–0.77, $P < 0.0001$) and the mean value was 3.44 ± 1.13 [0.00–7.19]. When the predictive score was larger than 4.60 (e.g., a female with BMI of 30 kg/m^2, minimum JSW of 1.90 mm, and osteophyte area of 5.00 mm^2) the chance of incident radiographic OA at T5y was 54% (PPV), which was evidently larger than the incidence of radiographic OA in 19% of all knees (prior probability; 189 of 985 knees had 'poor' outcome). The chance of not developing radiographic OA (NPV) was 93% instead of 81% (prior probability) when the predictive score was 2.50 or lower (e.g., a female with BMI of 30 kg/m^2, minimum JSW of 3.00 mm and osteophyte area of 0.00 mm^2).

Clinical OA: persistence

Table V summarizes results of regression analyses on persistent knee pain. The difference-values of the radiographic characteristics (knee − contralateral knee) are not depicted since none of these variables were significant predictors of this outcome in univariate and multivariate analyses.

The predictive value (OR) of radiographic characteristics was smaller for persistent knee pain outcome than for radiographic outcome in univariate and multivariate analyses. Of the participants with radiographic OA at T5y, 53% also had persistent pain (95 of 179). And of the participants with persistent pain, 22% had radiographic OA at T5y.

Hosmer–Lemeshow tests showed adequate fit for the final models with radiographic characteristics (KIDA and K&L: $P > 0.05$), but lack of fit for the model with demographics only ($P = 0.01$).

The multivariate models all implied poor prognostic ability. The model with demographic and clinical characteristics only ($n = 1035$ of 1060 knees) had AUC–ROC 0.58 (95%CI: 0.54–0.61, $P < 0.001$). Adding radiographic variables hardly improved the ability to predict pain persistence: AUC–ROC = 0.60 (95%CI: 0.56–0.64, $P < 0.001$) for the separate radiographic features ($n = 957$ knees) and AUC–ROC = 0.60 (95%CI: 0.56–0.63, $P < 0.01$) for K&L-grade ($n = 970$ knees).

Clinical OA: progression

Tables VIA and VIB depict results of regression analyses with WOMAC pain and function outcome as dependent variable, respectively. The multivariate model with demographics and clinical characteristics ($n = 282$ and 267 participants, respectively) had AUC–ROC of 0.59 (95%CI: 0.52–0.66, $P < 0.01$) and 0.63 (95%CI: 0.56–0.70, $P < 0.001$), respectively. Adding radiographic features slightly improved the prediction of WOMAC pain and function development [AUC–ROC = 0.62 (95%CI: 0.55–0.68, $P < 0.001$) and 0.65 (95%CI:0.58–0.71, $P < 0.001$), for pain and function, respectively]. Interestingly, in the multivariate model for WOMAC pain the difference-value (between contralateral knees) of eminence height was a significant predictor of 'poor' outcome. When adding K&L-grade (difference) to demographics and clinical variables comparable poor predictive abilities were found.

Hosmer–Lemeshow tests revealed adequate fit for all multivariate models with WOMAC outcome. For the clinical outcomes, the AUC–ROC was not statistically significantly different between the multivariate models [e.g., model with demographics and clinical variables only versus model where radiographic variables (features or K&L-grade) were added].

For the analyses on participant level a portion of the individuals contributed with both knees to the regression analyses. To account for this dependency generalized estimating equations (GEE) were performed, which resulted in (nearly) the same OR and *P*-values as the regression analyses.

Table V
Regression analyses with persistent knee pain as dependent variable

Univariate				Multivariate				
	OR	(95%CI)	*P*		OR	(95%CI)	*P*	AUC–ROC
Demographic & clinical				**Demographic & clinical**				0.58
Age	0.97	(0.95–1.00)	0.03					
Gender	0.99	(0.73–1.35)	0.95					
BMI	1.03	(1.00–1.06)	0.03					
ESR	1.00	(0.98–1.01)	0.58					
Pain intensity	1.15	(1.08–1.22)	<0.0001	Pain intensity	1.15	(1.08–1.22)	<0.0001	
Radiographic				**Radiographic feature (demographic & clinical)**				
Key feature				***Key feature***				0.60
Minimum JSW	0.88	(0.80–0.98)	0.02	Minimum JSW	0.88	(0.79–0.97)	0.01	
Medial JSW	0.91	(0.80–1.04)	0.16					
Lateral JSW	0.96	(0.88–1.05)	0.37					
Varus angle	1.00	(0.94–1.08)	0.90					
Osteophyte	1.48	(1.09–2.02)	0.01	Osteophyte	1.54	(1.12–2.13)	0.008	
Eminence	1.01	(0.97–1.05)	0.73					
Bone density	1.02	(0.99–1.04)	0.21					
				Age	0.97	(0.94–0.99)	0.01	
				Pain intensity	1.13	(1.06–1.20)	0.0002	
				K&L (demographic & clinical)				
K&L-grade	1.42	(1.12–1.82)	0.004	***K&L-grade***	1.47	(1.14–1.88)	0.002	0.60
				Age	0.97	(0.95–1.00)	<0.0001	
				Pain intensity	1.13	(1.07–1.21)	0.04	

P: significance level; AUC–ROC: area under the receiver operating characteristic curve; note: pain intensity (and age) depicted for three multivariate models: demographic & clinical, and key feature and K&L-grade in addition to demographic & clinical.

Table VIA
Regression analyses with WOMAC pain outcome as dependent variable

Univariate				Multivariate				
	OR	(95%CI)	*P*		OR	(95%CI)	*P*	AUC–ROC
Demographic & clinical				**Demographic & clinical**				0.59
Age	0.96	(0.92–1.01)	0.13					
Female gender	1.40	(0.80–2.46)	0.24					
BMI	1.09	(1.02–1.17)	0.01	BMI	1.09	(1.02–1.17)	0.01	
ESR	1.05	(1.01–1.09)	0.02					
WOMAC pain	1.02	(1.00–1.03)	0.02					
Radiographic				**Radiographic feature (demographic & clinical)**				
Key feature				***Key feature***				0.62
Minimum JSW	0.98	(0.88–1.09)	0.77					
Medial JSW	1.09	(0.97–1.23)	0.16					
Lateral JSW	1.01	(0.92–1.11)	0.82					
Varus angle	0.97	(0.91–1.04)	0.46					
Osteophyte	1.37	(0.76–2.46)	0.29					
Eminence	1.01	(0.97–1.05)	0.55					
Bone density	0.99	(0.97–1.01)	0.24					
Key feature abs diff				***Key feature abs diff***				
Minimum JSW	1.13	(0.82–1.57)	0.45					
Medial JSW	1.05	(0.69–1.60)	0.81					
Lateral JSW	0.94	(0.73–1.21)	0.64					
Varus angle	0.94	(0.80–1.11)	0.49					
Osteophyte	1.44	(0.75–2.74)	0.28					
Eminence	0.86	(0.73–1.03)	0.10	Eminence	0.81	(0.67–0.98)	0.03	
Bone density	0.99	(0.86–1.13)	0.89					
				ESR	1.05	(1.01–1.09)	0.02	
				WOMAC pain	1.02	(1.00–1.03)	0.05	
K&L-grade	0.97	(0.76–1.25)	0.83	***K&L (demographic & clinical)***				
K&L abs diff	0.53	(0.29–0.98)	0.04	***K&L abs diff***	0.46	(0.24–0.87)	0.02	0.63
				ESR	1.05	(1.01–1.09)	0.03	
				WOMAC pain	1.02	(1.00–1.04)	0.03	

P: significance level; AUC–ROC: area under the receiver operating characteristic curve; abs diff: absolute difference; note: ESR and WOMAC pain depicted for two multivariate models: key feature and K&L-grade in addition to demographic & clinical.

Table VIB
Regression analyses with WOMAC function outcome as dependent variable

Univariate				Multivariate				
	OR	(95%CI)	*P*		OR	(95%CI)	*P*	AUC–ROC
Demographic & clinical				**Demographic & clinical**				0.63
Age	1.01	(0.96–1.06)	0.74					
Female gender	1.79	(1.01–3.18)	0.05					
BMI	1.09	(1.02–1.16)	0.02					
ESR	1.06	(1.02–1.11)	0.002	ESR	1.05	(1.01–1.10)	0.01	
WOMAC function	1.02	(1.00–1.04)	0.01	WOMAC function	1.02	(1.00–1.04)	0.02	
Radiographic				**Radiographic feature (demographic & clinical)**				
Feature				***Feature***				0.65
Minimum JSW	0.89	(0.80–0.99)	0.03	Minimum JSW	0.89	(0.80–0.99)	0.04	
Medial JSW	0.89	(0.79–1.01)	0.07					
Lateral JSW	0.99	(0.90–1.09)	0.79					
Varus angle	1.04	(0.97–1.11)	0.31					
Osteophyte	1.31	(0.72–2.38)	0.38					
Eminence	1.00	(0.96–1.04)	0.83					
Bone density	0.99	(0.97–1.00)	0.10					
Feature abs diff				***Feature abs diff***				
Minimum JSW	1.04	(0.75–1.45)	0.82					
Medial JSW	1.18	(0.75–1.86)	0.47					
Lateral JSW	1.00	(0.77–1.29)	0.97					
Varus angle	1.02	(0.87–1.23)	0.75					
Osteophyte	1.95	(1.00–3.82)	0.05					
Eminence	0.91	(0.76–1.08)	0.27					
Bone density	0.96	(0.83–1.10)	0.54					
				ESR	1.05	(1.01–1.09)	0.02	
				WOMAC function	1.02	(1.00–1.04)	0.02	
K&L-grade	1.10	(0.85–1.42)	0.46					
K&L abs diff	0.71	(0.39–1.30)	0.71					

P: significance level; AUC–ROC: area under the receiver operating characteristic curve; abs diff: absolute difference; note: ESR and WOMAC pain depicted for the multivariate models with radiographic features in addition to demographic & clinical.

Discussion

In individuals that presented themselves with very early complaints related to knee OA, evaluation of radiographic characteristics added to the prediction of incident radiographic OA 5 years later. The evaluation of separate quantitative features performed better in this respect than a simple K&L-grade. Radiographic characteristics hardly added to the prediction of persistence and/or progression of clinical OA, and total predictive ability of these models was too low for use in practice.

The additional value of radiographic characteristics in the prediction of radiographic progression has been described for more advanced OA (e.g., K&L-grade $\geq$II)[14,29–31]. And in a recent review also radiographic features, varus alignment, age, and BMI[32] were identified as predictors for OA progression later in disease. This is the first study however to demonstrate that quantitative radiographic features, identified in individuals that present themselves with knee pain but without radiographic damage (K&L-grade <II), can add to the prediction of incident radiographic OA within 5 years.

The finding that radiographic characteristics were of additional value in the prediction of radiographic outcome, but hardly in the prediction of clinical outcome[18,32] is in accordance with the commonly reported inconsistent association between radiographic and clinical characteristics of OA[8,33]. Even when structural damage was evaluated with magnetic resonance imaging, only a weak correlation between change in WOMAC score and cartilage thickness was found[34].

In the present study basic demographics together with clinical characteristics poorly predicted clinical outcome (persistence and/or development). This limited predictive ability for clinical outcome might be explained by the subjective nature of these outcomes. A limited set of variables that was commonly used in patient care was chosen for analyses, based on undemanding application in clinical practice.

Despite the limited severity and development of complaints in this very early OA cohort, the quintile approach[24] discriminated participants with 'good' and 'poor' outcome with significantly different scores at T5y. The WOMAC function score was 6 [1–12] for participants with 'good' and 29 [19–41] for participants with 'poor' outcome, and the WOMAC pain score was 8 [5–15] for those with 'good' outcome and 28 [20–40] for those with 'poor' outcome. Irrespectively, specifically these individuals [at risk of developing (radiographic) OA] present themselves with complaints for the first time. And although at that time radiography (and demographic and clinical characteristics) can hardly predict clinical outcome, specific key features obtained with KIDA measurement significantly add to the prediction of radiographic outcome.

The fact that WOMAC outcomes were measured on participant level, contrary to the radiographic outcome on joint level, might also in part explain the limited predictive ability of radiographic characteristics for this clinical outcome. The sum of the radiographic characteristics might be more appropriate to detect an association with clinical outcome in case of (more severe) bilateral OA, when the radiographic characteristics are expected to be more pronounced. In case of (milder) unilateral OA, as will be the case when individuals present themselves for the first time with complaints, the sum of one unaffected and one (slightly) affected joint might underestimate radiographic severity and the difference between the knees might be more appropriate[25]. In general however, in our study difference-values between both knee joints in the radiographic characteristics did not appear better in predicting the incidence of radiographic OA or the persistence or progression of clinical OA. Surprisingly, the difference-values between both knees in eminence height and in K&L-grade were found to be predictors of WOMAC pain outcome, while the sum of the measured values was not significantly associated with this outcome. The ORs were low however, and moreover both OR's were smaller than 1 which implies that a larger difference between joints is protective for 'poor' outcome in contrast to our hypothesis. However, the predictive ability of the models (AUC–ROC) was poor indicating that predicting WOMAC outcome is difficult either with or without radiographic features.

The separate key features that were identified as additional predictors for incidence of radiographic OA and for clinical persistence were minimum JSW and osteophyte area. These separate features measured quantitatively using KIDA are also the most important characteristics in K&L-grading. This explains why also K&L-grading added to demographics and clinical variables in predicting radiographic incidence of OA. When all radiographic variables (minimum JSW, osteophyte area and K&L-grading) were added in logistic regression analysis, these were all significant predictors and predictive ability was even slightly higher than for KIDA variables and K&L-grading separately (AUC–ROC = 0.76). Since OR was strongest for osteophyte area (5.02), measuring separate features is of value in addition to demographic and clinical characteristics (and K&L-grading). Also, since the KIDA predictors performed better than K&L-grading in this cohort, measuring separate radiographic features might improve the detection of radiographic OA earlier in the disease process. This was supported by the detection of joint space narrowing and osteophyte formation when qualitatively evaluating separate parameters by the Altman atlas[35], with larger AUC–ROC than K&L-grade (0.76, data not shown).

Next to the advantage of evaluation of separate key features of joint damage, KIDA measurement uses a mathematical approach and is performed without any knowledge of the knee and the individual. Intra-observer variation of KIDA was low[22,25], however variation in the measurements might occur during image acquisition. Despite optimal standardization[36] the position of the tibial plateau is subject to variation[37], which decreases comparability between and within individuals. Due to such variations, the additional value of KIDA might be underestimated in the present study, and might be improved when reproducibility of radiographic acquisition is further optimized in clinical trials. Also the method of measuring the different joint characteristics might be optimized by comparison with other available techniques, such as different methods for varus angle measurement with small minimal detectable differences[38]. Further, the predictive ability of K&L-grading might be overestimated since in the present study the definition of K&L-grade at T0 was determined with knowledge of the K&L-grade at T2y (although not at T5y, this might be regarded a 'proxy' and as such the T0 measurement can not be regarded fully blinded like the KIDA measurements). This is of course also contrary to clinical practice when individuals present themselves with joint pain related to OA, and a choice on treatment (or on inclusion in a trial) is preferably made within a short time span, without waiting for a second radiograph 1 or 2 years later.

By use of the predictive scores (based on separate features) a subgroup of individuals was distinguished with a higher chance of onset of radiographic OA (54% compared to incidence of 19%). Although this chance is too low for decision-making at the individual level, identifying this group might advance the design of OA trials[39] for the development of more specific (disease modifying) treatment strategies[40]. The predictive score and cut-off values as determined in this study need internal and external validation before used in (clinical) practice. Also the fit of the models might be improved by investigating for instance different transformations of the predictors.

In conclusion, the prediction of incidence of radiographic OA improves from poor to fair when quantitative radiographic features are evaluated, in addition to basic demographics and clinical assessment, in individuals that visit a physician with early complaints possibly related to OA. Therefore the measurement of separate features might be valuable in identifying individuals at high risk of developing radiographic OA.

Author contributions

MK, AM, MV, JB, PW, and FL contributed to conception and design of this study. KD and SB added to data collection. MK, KV and PW contributed to data analyses and MK, AM, PW, and FL contributed to interpretation of data. Article drafts were written by MK and critically revised by all authors. The final version of the article was approved by all the authors.

Conflict of interest

The authors declare that they have no conflicts of interest.

Acknowledgements

We acknowledge Marja Lafeber (ML) for KIDA analyses. CHECK is funded by the Dutch Arthritis Association.

References

1. Bijlsma JW, Berenbaum F, Lafeber FP. Osteoarthritis: an update with relevance for clinical practice. Lancet 2011 Jun 18;377(9783):2115–26.
2. Peat G, Thomas E, Duncan R, Wood L, Hay E, Croft P. Clinical classification criteria for knee osteoarthritis: performance in the general population and primary care. Ann Rheum Dis 2006 Oct;65(10):1363–7.
3. Nevitt MC. Definition of hip osteoarthritis for epidemiological studies. Ann Rheum Dis 1996 Sep;55(9):652–5.
4. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. Ann Rheum Dis 1957 Dec;16(4):494–502.
5. Boegard TL, Rudling O, Petersson IF, Jonsson K. Joint space width of the tibiofemoral and of the patellofemoral joint in chronic knee pain with or without radiographic osteoarthritis: a 2-year follow-up. Osteoarthritis Cartilage 2003 May;11(5):370–6.
6. Vignon E, Conrozier T, Hellio Le Graverand MP. Advances in radiographic imaging of progression of hip and knee osteoarthritis. J Rheumatol 2005 Jun;32(6):1143–5.
7. Dieppe PA, Cushnaghan J, Shepstone L. The Bristol 'OA500' study: progression of osteoarthritis (OA) over 3 years and the relationship between clinical and radiographic changes at the knee joint. Osteoarthritis Cartilage 1997 Mar;5(2):87–97.
8. Kinds MB, Welsing PM, Vignon EP, Bijlsma JW, Viergever MA, Marijnissen AC, *et al*. A systematic review of the association between radiographic and clinical osteoarthritis of hip and knee. Osteoarthritis Cartilage 2011 Jan 31.
9. Hunter DJ. Lower extremity osteoarthritis management needs a paradigm shift. Br J Sports Med 2011 Feb 6.
10. Felson DT, Zhang Y. An update on the epidemiology of knee and hip osteoarthritis with a view to prevention. Arthritis Rheum 1998 Aug;41(8):1343–55.
11. Grotle M, Hagen KB, Natvig B, Dahl FA, Kvien TK. Obesity and osteoarthritis in knee, hip and/or hand: an epidemiological study in the general population with 10 years follow-up. BMC Musculoskelet Disord 2008;9:132.
12. Heliovaara M, Makela M, Impivaara O, Knekt P, Aromaa A, Sievers K. Association of overweight, trauma and workload with coxarthrosis. A health survey of 7,217 persons. Acta Orthop Scand 1993 Oct;64(5):513–8.
13. Ding C, Jones G, Wluka AE, Cicuttini F. What can we learn about osteoarthritis by studying a healthy person against a person with early onset of disease? Curr Opin Rheumatol 2010 Sep;22(5):520–7.
14. Eckstein F, Le Graverand MP, Charles HC, Hunter DJ, Kraus VB, Sunyer T, *et al*. Clinical, radiographic, molecular and MRI-based predictors of cartilage loss in knee osteoarthritis. Ann Rheum Dis 2011 Jul;70(7):1223–30.
15. Dougados M, Gueguen A, Nguyen M, Berdah L, Lequesne M, Mazieres B, *et al*. Requirement for total hip arthroplasty: an outcome measure of hip osteoarthritis? J Rheumatol 1999 Apr;26(4):855–61.
16. Vinciguerra C, Gueguen A, Revel M, Heuleu JN, Amor B, Dougados M. Predictors of the need for total hip replacement in patients with osteoarthritis of the hip. Rev Rhum Engl Ed 1995 Oct;62(9):563–70.
17. Belo JN, Berger MY, Reijman M, Koes BW, Bierma-Zeinstra SM. Prognostic factors of progression of osteoarthritis of the knee: a systematic review of observational studies. Arthritis Rheum 2007 Feb 15;57(1):13–26.
18. Thomas E, Peat G, Mallen C, Wood L, Lacey R, Duncan R, *et al*. Predicting the course of functional limitation among older adults with knee pain: do local signs, symptoms and radiographs add anything to general indicators? Ann Rheum Dis 2008 Oct;67(10):1390–8.
19. Wesseling J, Dekker J, van den Berg WB, Bierma-Zeinstra SM, Boers M, Cats HA, *et al*. CHECK (Cohort Hip and Cohort Knee): similarities and differences with the Osteoarthritis Initiative. Ann Rheum Dis 2009 Sep;68(9):1413–9.
20. Buckland-Wright C. Protocols for precise radio-anatomical positioning of the tibiofemoral and patellofemoral compartments of the knee. Osteoarthritis Cartilage 1995 Sep;3(Suppl A):71–80.
21. Buckland-Wright JC, Wolfe F, Ward RJ, Flowers N, Hayne C. Substantial superiority of semiflexed (MTP) views in knee osteoarthritis: a comparative radiographic study, without fluoroscopy, of standing extended, semiflexed (MTP), and schuss views. J Rheumatol 1999 Dec;26(12):2664–74.
22. Marijnissen AC, Vincken KL, Vos PA, Saris DB, Viergever MA, Bijlsma JW, *et al*. Knee Images Digital Analysis (KIDA): a novel method to quantify individual radiographic features of knee osteoarthritis in detail. Osteoarthritis Cartilage 2008 Feb;16(2):234–43.
23. van Dijk GM, Veenhof C, Spreeuwenberg P, Coene N, Burger BJ, van Schaardenburg D, *et al*. Prognosis of limitations in activities in osteoarthritis of the hip or knee: a 3-year cohort study. Arch Phys Med Rehabil 2010 Jan;91(1):58–66.
24. Sharma L, Cahue S, Song J, Hayes K, Pai YC, Dunlop D. Physical functioning over three years in knee osteoarthritis: role of psychosocial, local mechanical, and neuromuscular factors. Arthritis Rheum 2003 Dec;48(12):3359–70.
25. Kinds MB, Vincken KL, Vignon E, ten Wolde S, Bijlsma JW, Welsing PM, *et al*. Radiographic features of knee and hip osteoarthritis represent characteristics of an individual, in addition to severity of osteoarthritis. Scand J Rheumatol, submitted for publication. Epub 16 Dec 2011.
26. Muller MP, Tomlinson G, Marrie TJ, Tang P, McGeer A, Low DE, *et al*. Can routine laboratory tests discriminate between severe acute respiratory syndrome and other causes of community-acquired pneumonia? Clin Infect Dis 2005 Apr 15;40(8):1079–86.
27. Van Houwelingen JC, Le Cessie S. Predictive value of statistical models. Stat Med 1990 Nov;9(11):1303–25.

28. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology 1983 Sep;148(3):839–43.
29. Mazzuca SA, Brandt KD, Katz BP, Ding Y, Lane KA, Buckwalter KA. Risk factors for progression of tibiofemoral osteoarthritis: an analysis based on fluoroscopically standardised knee radiography. Ann Rheum Dis 2006 Apr; 65(4):515–9.
30. Bruyere O, Honore A, Ethgen O, Rovati LC, Giacovelli G, Henrotin YE, *et al*. Correlation between radiographic severity of knee osteoarthritis and future disease progression. Results from a 3-year prospective, placebo-controlled study evaluating the effect of glucosamine sulfate. Osteoarthritis Cartilage 2003 Jan;11(1):1–5.
31. Wolfe F, Lane NE. The longterm outcome of osteoarthritis: rates and predictors of joint space narrowing in symptomatic patients with knee osteoarthritis. J Rheumatol 2002 Jan;29(1):139–46.
32. Chapple CM, Nicholson H, Baxter GD, Abbott JH. Patient characteristics that predict progression of knee osteoarthritis: a systematic review of prognostic studies. Arthritis Care Res (Hoboken) 2011 Aug;63(8):1115–25.
33. Bedson J, Croft PR. The discordance between clinical and radiographic knee osteoarthritis: a systematic search and summary of the literature. BMC Musculoskelet Disord 2008;9:116.
34. Buck R, Nevitt M, Wirth W, Eckstein F. Correlation of knee cartilage thickness change at 1 year and clinical outcome changes at 3 years: data from the osteoarthritis initiative. Presented at OARSI imaging workshop. 2011.
35. Altman RD, Gold GE. Atlas of individual radiographic features in osteoarthritis, revised. Osteoarthritis Cartilage 2007; 15(Suppl A):A1–A56.
36. Mazzuca SA, Brandt KD, Buckwalter KA, Lane KA, Katz BP. Field test of the reproducibility of the semiflexed metatarsophalangeal view in repeated radiographic examinations of subjects with osteoarthritis of the knee. Arthritis Rheum 2002 Jan;46(1):109–13.
37. Botha-Scheepers S, Dougados M, Ravaud P, Hellio Le Graverand MP, Watt I, Breedveld FC, *et al*. Effect of medial tibial plateau alignment on serial radiographs on the capacity to predict progression of knee osteoarthritis. Osteoarthritis Cartilage 2008 Feb;16(2):272–6.
38. McDaniel G, Mitchell KL, Charles C, Kraus VB. A comparison of five approaches to measurement of anatomic knee alignment from radiographs. Osteoarthritis Cartilage 2010 Feb;18(2):273–7.
39. Lane NE, Brandt K, Hawker G, Peeva E, Schreyer E, Tsuji W, *et al*. OARSI-FDA initiative: defining the disease state of osteoarthritis. Osteoarthritis Cartilage 2011 May;19(5):478–82.
40. Felson DT. Identifying different osteoarthritis phenotypes through epidemiology. Osteoarthritis Cartilage 2010 May; 18(5):601–4.